Attention

HE Jiayou

July 5, 2022

Table of Contents

Notations

n: Batchsize

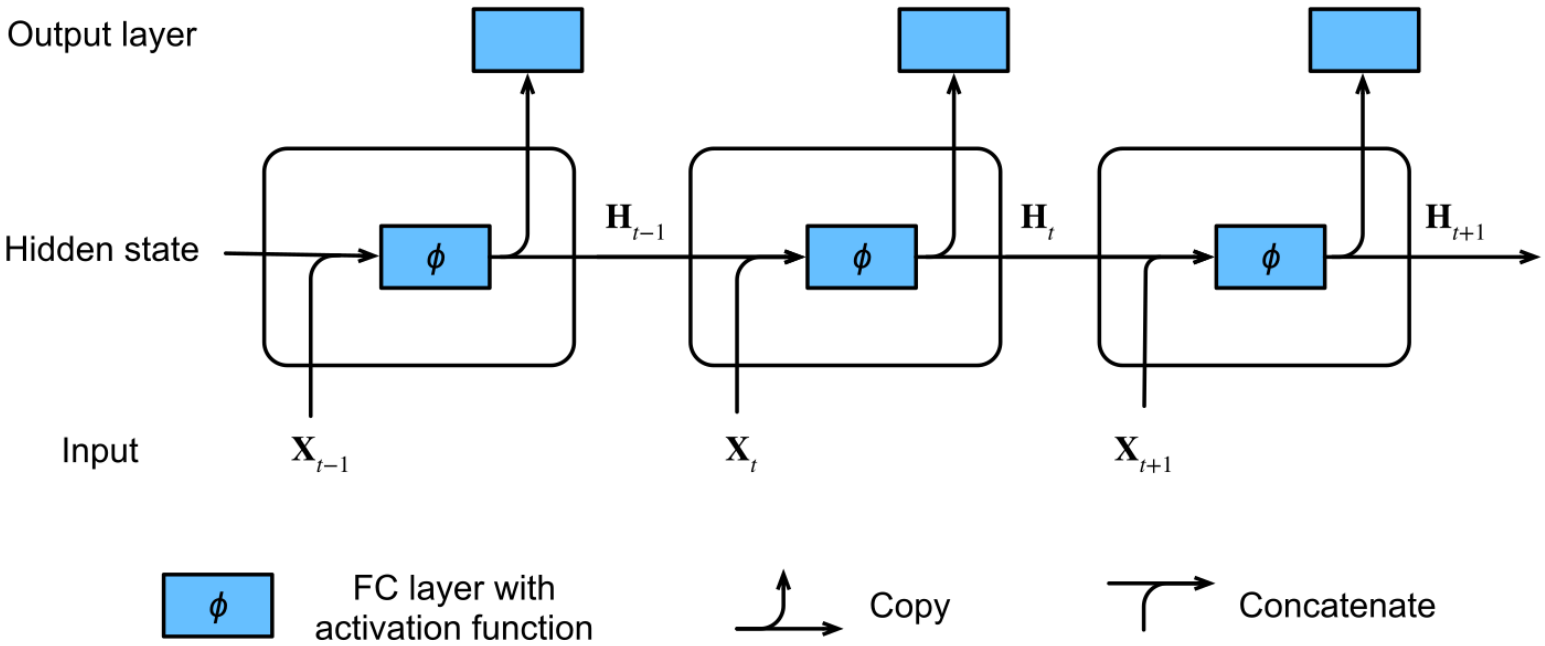
d: Dimension of input

h: Dimension of hidden state

q: Dimension of answer

Table of Contents

RNN



$$H_t = \phi\left(X_t W_{xh} + H_{t-1} W_{hh} + b_h\right)$$

$$O_t = H_t W_{hq} + b_q$$

Table of Contents

GRU Gated Recurrent Unit

$$
\begin{aligned}
R_t &= \sigma\left(X_t W_{xr} + H_{t-1} W_{hr} + b_r\right) \\
Z_t &= \sigma\left(X_t W_{xz} + H_{t-1} W_{hz} + b_z\right) \\
\tilde{H}_t &= \tanh\left(X_t W_{xh} + \left(R_t \odot H_{t-1}\right) W_{hh} + b_h\right) \\
H_t &= Z_t \odot H_{t-1} + \left(1 - Z_t\right) \odot \tilde{H}_t
\end{aligned}
$$

Being as an encoder.

Table of Contents

LSTM

Similar to GRU

$$
\begin{aligned}
I_t &= \sigma\big(X_t W_{xi} + H_{t-1} W_{hi} + b_i\big) \\
F_t &= \sigma\big(X_t W_{xf} + H_{t-1} W_{hf} + b_f\big) \\
O_t &= \sigma\big(X_t W_{xo} + H_{t-1} W_{ho} + b_o\big) \\
\tilde{C}_t &= \tanh\big(X_t W_{xc} + H_{t-1} W_{hc} + b_c\big) \\
C_t &= F_t \odot C_{t-1} + I_t \odot \tilde{C}_t \\
H_t &= O_t \odot \tanh(C_t)
\end{aligned}
$$

Table of Contents

Encoder-Decoder

Table of Contents

Attention Prompt

Table of Contents

Transformer

Table of Contents

architecture

Table of Contents

Attention

Table of Contents

Scoring Function

Table of Contents

Multi-Head Attention

Table of Contents

Self-Attention

Table of Contents

Positional Encoding

Table of Contents

Residual Connection & Layer Normalization

Table of Contents

Medical-Related

Table of Contents

Attention UNet

Table of Contents

Multi-scale Self-guided Attention